Preventive Veterinary Medicine 101 (2011) 35–41

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Analysis of the diagnostic accuracy of the gamma interferon assay for detection of bovine tuberculosis in U.S. herds

María C. Antognoli[a,*], Marta D. Remmenga[a], Steven D. Bengtson[a], Halden J. Clark[a], Kathy A. Orloski[b], Lori L. Gustafson[a], Aaron E. Scott[a]

[a] USDA-APHIS-VS-CEAH, National Surveillance Unit, 2150 Center Avenue Building B, MS 2E6, Fort Collins, CO, 80526-8117, USA
[b] USDA-APHIS-VS National Tuberculosis Eradication Program, Fort Collins, CO, USA

ARTICLE INFO

Article history:
Received 10 January 2011
Received in revised form 17 May 2011
Accepted 19 May 2011

Keywords:
Bovine tuberculosis
Evaluation of diagnostic tests
Gamma interferon assay
Mycobacterium bovis

ABSTRACT

The goal of this study was to evaluate the test sensitivity (SE) and specificity (SP) of the gamma interferon (G-IFN) assay used for the detection of bovine tuberculosis (bTB) in U.S. cattle herds. In addition, the study assessed the association between G-IFN test results and bTB status of cattle, and explored different cut off values for classification of test results in adult cattle using receiver operating characteristics (ROC) curve analysis.

Test SE was estimated using a population of 87 confirmed infected cattle from 14 herds distributed in 6 states. Test SP was estimated using a population of 4123 cattle representing 3000 premises in 3 states. These animals were from bTB free areas, accredited bTB free herds, or herds that were historically bTB free based on the absence of lesions found at slaughter and historical records of negative tests performed for bTB surveillance. The distribution of G-IFN results and its association with bTB infection status was also explored in a group of 914 exposed cattle in which infection was not confirmed.

The results showed that the SE of the G-IFN for a cut-off value ≥0.1 was 83.9% (76.1, 91.6). The SP of the G-IFN was 90.7% (95% CI: 89.8, 91.6), 97% (95% CI: 96.5, 97.5), and 98.6%(95% CI: 98.2, 98.9), for cut off values of 0.1, 0.3, and 0.5, respectively. For a cut off value ≥0.1, the likelihood ratio of a positive G-IFN test was 9.03 (95% CI: 7.90, 10.31), and the likelihood ratio of a negative G-IFN test was 0.18 (95% CI: 0.11, 0.29).

The area under the ROC curve was 0.976 (95% CI: 0.97, 0.98), characteristic of a highly accurate test. ROC analysis also showed that lower cut-off values, such as 0.1, have high SE with suitable SP for use in parallel testing, while cut-off values ranging between 0.3 and 0.6 provide the high SP desired in series-testing protocols with lower SE values.

Findings from this study indicated that the G-IFN performs with high accuracy in the field, yielding SE and SP estimates comparable to those reported in previous evaluations (Ryan et al., 2000; Ameni et al., 2000; de la Rua-Domenech et al., 2006; Gormley et al., 2006).

Published by Elsevier B.V.

1. Introduction

The gamma interferon (G-IFN) is an *in vitro* assay that detects a cell-mediated immune response. This assay is conducted by incubation of blood with *Mycobacterium bovis* purified protein derivative (PPD) and *M. avium* PPD, followed by use of an enzyme-linked immunosorbent

* Corresponding author. Tel.: +1 970 494 7304; fax: +1 970 494 7174.
E-mail address: Celia.Antognoli@aphis.usda.gov (M.C. Antognoli).

0167-5877/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.prevetmed.2011.05.012

assay (ELISA) to detect and quantify the release of the cytokine gamma interferon by lymphocytes from infected animals (Rothel et al., 1990). Sensitized T-lymphocytes from infected animals recognize specific mycobacterial antigens from *M. bovis* and are stimulated to secrete G-IFN, while those from uninfected cattle not previously exposed are not expected to respond (Cagiola et al., 2004).

Among the advantages of the G-IFN over the intradermal tuberculin tests are less frequent handling of animals, objective numerical test results, comparative response to two different antigens, and the capability of adjusting the sensitivity (SE) and specificity (SP) of the test with different cut-off values. Conversely, the G-IFN depends on the presence of viable lymphocytes in the blood sample collected in the field. Careful handling of blood samples and timely transportation to the laboratory are critical to this test and are limitations of this diagnostic tool.

In 2002, the TB Scientific Advisory Subcommittee (TB-SAS) of the United States Animal Health Association (USAHA) recommended the use of the G-IFN test (Bovigam™) as an ancillary or supplemental test in TB program herds known or suspected of having bovine TB (USAHA, 2002). This recommendation was incorporated in the United States Department of Agriculture (USDA) 2005 Bovine Tuberculosis Eradication Uniform Methods and Rules (UM&R) (USDA-APHIS, 2005). The UM&R specifies that the G-IFN test could be used: (a) in parallel testing with the Comparative Cervical Test (CCT), (b) in series with the CFT for retesting suspects (as a replacement for the CCT), and (c) in parallel with the CFT test or Cervical Test (CT) in affected herds. In practice, the G-IFN test is mostly used as a supplemental test in series with the CFT.

The accuracy of the G-IFN test has been the focus of several studies published in the United States and in other countries (Rothel et al., 1990; Wood et al., 1991, 1992; Neill et al., 1994; Whipple et al., 1995; Ameni et al., 2000; de la Rua-Domenech et al., 2006; Gormley et al., 2006). In these studies, however, the G-IFN test was used differently than it is currently used in the United States. First, the test procedures and its interpretation have changed over time: a control for lymphocyte viability (*in vitro* stimulation with pokeweed mitogen) was not described in the studies conducted in the early 1990s. Secondly, the test in the United States is currently performed 3–30 days after injection of the CFT. This prior in vivo stimulation with bovine PPD in the injected animals has proven to increase the SE of the test (Palmer et al., 2006). In the listed literature, the G-IFN was used as a primary test without prior in vivo stimulation with the CFT, which yields a different SE than the G-IFN as performed in the United States. These differences in testing protocol, interpretation of results, and test SE limit the ability to extrapolate results from prior publications to the present use of the G-IFN in the United States.

During fiscal year 2009, the USDA-Animal and Plant Health Inspection Services (APHIS)-Veterinary Services' (VS)-National Surveillance Unit (NSU) and TB program staff collaboratively conducted a new field performance analysis using testing data collected between 2005 and 2009. The objectives of this study were (1) to estimate diagnostic SE and SP of the G-IFN under field conditions, (2) to assess the association between G-IFN test results, bTB status of cattle, and postmortem test results in animals slaughtered due to bTB suspicion in the United States, and lastly (3) to explore the variability of the G-IFN optical densities and different cut-off points for classification of test results in adult cattle.

2. Methods

2.1. Study populations

The following data sets were used to conduct this evaluation:

Data set #1: This data set consisted of 1001 adult cattle, all of which had antemortem and postmortem testing results between 2005 and 2009. The time period was chosen to ensure that the G-IFN tests were performed and results were reported in a consistent and standardized format. These records were assembled from the National Veterinary Services Laboratories (NVSL) granuloma submission database, the Emergency Management Reporting System (EMRS), data submitted by individual designated TB epidemiologists (DTEs) in states where reactors had been found and slaughtered, and a few records from state Generic Databases (GDBs). These cattle belonged either to herds where bTB infection was confirmed or to herds that had received bTB exposed cattle from infected herds (traced out herds). The 1001 animals represented 56 premises distributed in six states: California, Colorado, Michigan, Minnesota, New Mexico, and Texas. The availability of complete testing results and the ability to link postmortem confirmation of bTB to animal ID and antemortem test results were the main criteria for inclusion of records in the data set. A data subset consisting of 87 cattle with confirmed bTB infection (24 beef and 53 dairy cattle) were used to estimate test SE. These infected cattle represented 14 premises distributed in six states, (CA, MI, MN, NM, CO and TX). The remaining animals in data set #1 (exposed, infection not confirmed) were used to describe the distribution of G-IFN test results in that group, and to assess associations between antemortem test results and animal infection status.

Data set #2: This data set was compiled exclusively from state GDBs and included G-IFN testing data from adult cattle that were tested for reasons other than bTB infection, bTB suspicion, or traceback investigations. The data set included 4123 records (3442 dairy cattle and the rest beef or mixed breeds) from three states (Texas = 1259, Michigan = 2105, and New Mexico = 759), from over 3000 premises for which G-IFN test OD values were available. These premises were assumed to be free from bTB based on a history of absence of bTB-like lesions at slaughter, negative antemortem test results conducted for the purposes of animal movement, herd accreditation/certification of bTB free status, routine bTB surveillance testing, milk ordinance, bTB free area testing, and a variety of other reasons unrelated to bTB suspicion. The exclusion of premises resulting from trace outs of bTB cases as well as those tested due to fence contact with an affected herd, provided further assurance that the animals originated in these premises were not exposed to bTB. Dataset #2 was used to estimate test SP and to develop a ROC curve.

2.1.1. Classification of herd bTB status

Herd status was established by VS regional TB epidemiologists and was based on criteria outlined in the 2005 UM&R: "Affected herd: A herd of livestock in which there is strong evidence that *Mycobacterium bovis* exists. This evidence should include, but is not limited to, any of the following: histopathology, polymerase chain reaction (PCR) assay, bacterial isolation or detection, testing data, or epidemiological evidence such as contact with known sources of infection" (USDA-APHIS, 2005).

2.1.2. Classification of individual animal bTB status

The infection status of individual animals included in data set #1 was determined as follows: (a) *Infected (I)*: Any animal that was bTB-compatible by histopathology and was confirmed bTB-infected by PCR *OR* any animal from which *M. bovis* was isolated by culture, and (b) *Exposed-Infection not confirmed (E-INC)*: Any animal subjected to necropsy due to suspicion of bTB infection resulting in absence of bTB gross lesions at postmortem, or presence of lesions that were found not compatible with bTB by histopathology and therefore were not subjected to confirmatory PCR testing.

Animals included in data set #2 were considered bTB-free based on the inclusion criteria used in selecting the herds. Herds included in this data set were tested for reasons other than bTB infection, bTB suspicion, or bTB traceback investigation. These animals were classified as not infected-not exposed (*NINE*).

2.1.3. G-IFN test

For all animals included in the study, the G-IFN test was applied between 3 and 30 days after a CFT (2005 UM&R). Results were classified by the testing laboratories based on the algorithm provided by the test manufacturer:

POSITIVE = (mean OD bovine PPD − mean OD nil antigen ≥ 0.1) AND (mean OD Bovine PPD − mean OD Avian PPD ≥0.1)
NEGATIVE = (mean OD bovine PPD − mean OD nil antigen < 0.1) OR (mean OD Bovine PPD − mean OD Avian PPD < 0.1)

In the remainder of this analysis, quantitative G-IFN test results and cut-off values are referred to as the OD value corresponding to the difference between mean OD bovine PPD minus the mean OD avian PPD.

2.2. Data analysis

2.2.1. Calculation of test SE, SP, and exact binomial 95% CIs

Test SE was calculated for 87 confirmed infected animals from data set#1 as the proportion of G-IFN test positive animals among bTB infected (definition criteria listed in the previous section).

For 72 out of the 87 infected animals used in the calculation of test SE, test results were expressed as OD values. For these records, test classification was based on a cut-off value of 0.1 for the difference between mean bovine and mean avian PPD OD values (i.e., positive if ≥0.1, negative if <0.1). This classification criterion is recommended by the kit manufacturer and used by accredited laboratories to report G-IFN test results.

Results of 15 out of 87 infected animals were expressed either as an OD range (i.e., 0.1 < bovine minus avian > 0.3) or as a categorical result (i.e., negative, positive). In both cases, final test classification was based on a cut-off value of ≥0.1 to remain consistent with the laboratories' reporting criteria.

Test SP was calculated by using data from data set#2. All records in this data set contained OD values. Three different cut-off values often used in the field were explored for the calculation of test SP: (a) a cut-off value ≥0.1, which categorizes a test as positive for any OD value ≥ 0.1 and as negative otherwise; (b) a cut-off value ≥0.3, indicating that a test is positive for any OD ≥ 0.3 and negative otherwise; and (c) a cut-off value ≥0.5, indicating that a test is positive for any OD ≥ 0.5 and negative otherwise. All herds included in the calculation of test SP had a history of negative CFT test results and an absence of cattle with bTB lesions at slaughter. All computations were performed using MedCalc Software (MedCalc version 11.3.6, Mariakerke, Belgium).

2.2.2. Association between G-IFN test results, bTB status of cattle, and postmortem test results

Likelihood ratios (LRs) were calculated to facilitate estimation of the post-test probability of bTB infection given a positive (LRP) or negative (LRN) G-IFN test result for bTB infected animals in data set 1 and for all animals included in dataset #2. Published guidelines indicate that test results with LRP > 10 or LRN < 0.1 produce substantial departure between the post-test and pre-test probability of disease compared with LRP between 5 and 10 or LRN between 0.1 and 0.2, which produce moderate changes in the post-test probability of disease. A LR of 1 indicates no change in the post-test probability of disease and signals a lack of test accuracy (Gardner and Greiner, 2006). The agreement between G-IFN test results and postmortem testing (histopathology) was assessed for records in data set 1 by calculating the observed and expected agreements as well as the agreement expected beyond chance (Kappa coefficient).[1]

2.2.3. Exploration of cut-off values

A receiver operating characteristic (ROC) curve was constructed using SAS ROCPLOT macro and MedCalc Software (MedCalc version 11.3.6, Mariakerke, Belgium) to examine the impact of different cut-off values on SE and SP, and to estimate the area under the curve (AUC). The AUC provides a global summary statistic of test accuracy and previously published guidelines suggest that 0.5 < AUC ≤ 0.7 represents low accuracy, 0.7 < AUC ≤ 0.9 moderate accuracy, and 0.9 < AUC ≤ 1.0 represents high accuracy (Swets, 1998). The ROC curve used to explore the impact of different cut-off

[1] Interpretation of Kappa: Poor agreement = <0.20; Fair agreement = 0.20–0.40 ;Moderate agreement = 0.40–0.60; Good agreement = 0.60–0.80; Very good agreement = 0.80–1.00.

Table 1
Number or herds and animals included in dataset 1, by State and by bTB status between 2005 and 2009.

State of origin	Herd TB status	Number of herds	Animals' TB status	Number of animals
California	Infected	5	I[†]	5
			E-INC	58
	Not infected	35	E-INC[*]	90
Subtotal		40		153
Colorado	Infected	1	I	1
			E-INC	8
Subtotal		1		9
Michigan	Infected	12	I	19
			E-INC	50
Subtotal		12		69
New Mexico	Infected	1	I	54
			E-INC	117
Subtotal		1		171
Minnesota	Infected	1	I	6
			E-INC	580
Subtotal		1		586
Texas	Infected	2	I	2
			E-INC	11
Subtotal		2		13
Totals		56		1001

[†] I, infected; E-INC, exposed infection not confirmed.
[*] These E-INC animals were in non-infected herds, but traced from an infected herd.

values on test SE and SP was created using data set #2 and a subset of infected animals from data set #1 ($n = 72$) for which quantitative OD values were available.

3. Results

3.1. Test sensitivity

The SE of the G-IFN in a data subset from dataset set #1 was 83.9% (76.1, 91.6). There were 87 bTB-infected animals in this data set, most located in New Mexico (54) and Michigan (18) and a few others in Minnesota (6), California (5), Texas (2) and Colorado (1). Test SE was computed only for a cut-off value ≥0.1 due to the small size of the infected cattle population in our data. Table 1 illustrates the contents and characteristics of data set #1 and Table 2 shows the results of the G-IFN test in each cattle population examined (I, E-INC, and NINE) for a cut off value ≥0.1.

3.2. Test specificity

The SP of G-IFN was computed using the 4123 records in data set #2. This data set included animals with no apparent exposure to bTB and OD values were given for every record. The larger size of this data set allowed the evaluation of different cut-off values for computation of test SP, resulting in increasing SP as the cut-off value increased. Thus, a cut-off value ≥0.1 resulted in a SP of 90.7% (89.8, 91.6), a cut-off value ≥0.3 gave SP of 97% (96.5, 97.5), and, lastly, a cut-off value ≥0.5 resulted in a SP of 98.6% (98.2, 98.9).

Table 2
Number of animals by G-IFN test results among infected, exposed-infection not confirmed, and not infected not exposed cattle (Data sets #1 and #2).

Test result	TB infected animals (I)	Exposed-infection not confirmed (E-INC)	Not infected not exposed (NINE)
G-IFN positive[*]	73	172	383
G-IFN negative[*]	14	742	3740
Total	87	914	4123

[*] Classification of G-IFN results is based on a cutoff value ≥0.1

3.3. Association between G-IFN test results, animal bTB status, and postmortem test results

3.3.1. Association between animal bTB status and G-IFN test results

The LR of a positive G-IFN test was 9.03 (95% CI: 7.90, 10.31) and the LR of a negative test was 0.18 (95% CI: 0.11, 0.29). The LRP indicates that the odds of disease are approximately 9 times the pre-test odds of disease if the G-IFN test result ≥0.1. The LRN indicates the odds of disease are one-fifth of the pre-test odds of disease if the G-IFN test result is <0.1.

3.3.2. Agreement between G-IFN and histopathology results

The percentages of observed and expected agreement between G-IFN results and histopathology were 96% and 94%, respectively. Kappa, the percentage of agreement expected beyond chance, was 39% (32.5, 46.1), which is considered fair agreement.

3.4. Distribution of G-IFN OD values among infected, exposed—infection not confirmed, and not infected not exposed cattle

The distributions of OD values for G-IFN test results among I, E-INC, and NINE cattle are shown in Fig. 1. Data

Table 3
G-IFN test cutoff values and their respective SE, SP, and LR values.

OD value†	SE	95% CI	SP	95% CI	LRP	95% CI	LRN	95% CI
>−0.002	98.6	92.5–100.0	74.6	73.3–76.0	3.8	3.8–4.0	0.0	0.003–0.1
>0.0905	94.3	86.4–98.5	89.0	88.0–89.9	8.6	8.1–9.1	0.1	0.02–0.2
>0.103	93.0	84.5–97.7	89.0	88.7–90.5	8.9	8.4–9.6	0.1	0.03–0.2
>0.254	84.7	74.3–92.1	95.9	95.3–96.5	20.9	19.0–23.1	0.2	0.09–0.3
>0.301	83.3	72.7–91.1	96.5	95.9–97.1	23.9	21.6–26.6	0.2	0.1–0.3
>0.394	79.1	68.0–87.8	97.6	97.2–98.1	34.2	30.4–38.6	0.2	0.1–0.3
>0.676	79.1	68.0–87.8	98.9	98.6–99.3	77.2	68.6–87.0	0.2	0.1–0.4

† Optical density values for the difference between Bovine minus Avian PPDs.

suggested that OD values for both not infected (E-INC and NINE) and infected animals were mostly clustered around mean and median OD values. However, for each group of animals (I, E-INC, and NINE) there was a wide range of dispersed OD values that challenge the selection of a cut-off for classification of test results.

3.5. ROC curve analysis

Fig. 2 details the critical area of a ROC curve for G-IFN results from 75 records from bTB infected animals in data set # 1 (with quantitative G-IFN test results) and from all records in data set #2. The AUC was 0.976 (0.97, 0.98) with a standard error of 0.006. Based on guidelines detailed in the methodology section, the value of the AUC suggests that the G-IFN, as used under field conditions, is a highly accurate test. Table 3 shows the variation in SE, SP, LRN and LRP for some of the cut-off values most commonly used by field epidemiologists to classify animal status based on the G-IFN test. For example, a cut-off value ≥0.1 provided a SE of 93% (84.3, 97.7) and a SP of 89% (88.4, 90.1).

At a cut-off value ≥0.301, the SE decreased to 83% (72.3, 91.0) and the SP increased to 97% (96.8, 97.7), and at a cut-off value of 0.676 the SE decreased even more to 78.9% (67.6, 87.7) and the SP increased to 98.5% (98.2, 98.8). Table 3 also shows that the 95% confidence limits associated with SE estimates were wider than those of the SP estimates. The difference in the width of the 95% confidence intervals between SE and SP estimates could be attributed to the small sample size of bTB infected cattle with G-IFN results ranging between 0.1 and 0.6.

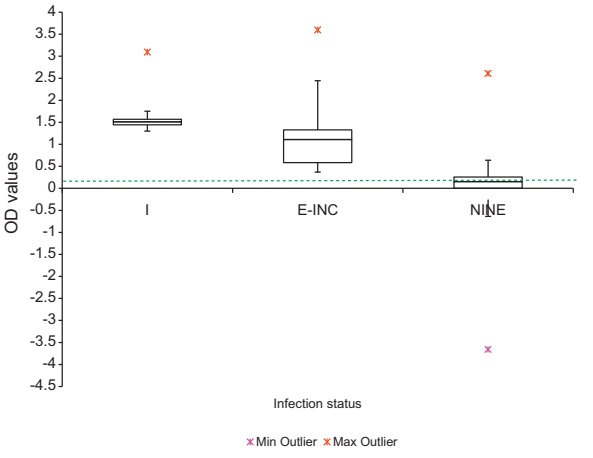

Fig. 1. Distribution of OD values for G-IFN test results among infected (I), exposed-infection not confirmed (E-INC) cattle, and not infected not exposed (NINE) cattle. The upper lines in the boxes represent the 75th percentile (P75), the middle line represents the median (P50), and the lower line in the box represents the 25th percentile (P25). The ends of the whiskers represent minimum and maximum OD values. The dashed horizontal line at the 0.10 OD value indicates the cut off value used for classification of G-IFN results in this study.

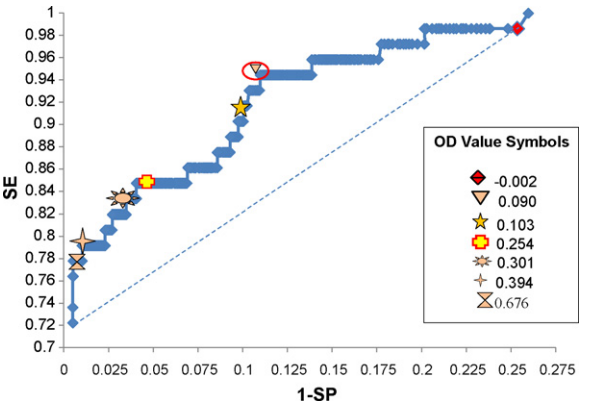

Fig. 2. Detail of the upper left corner of a ROC curve for G-IFN OD values. The different symbols represent G-IFN OD values at each combination of SE and 1-SP. The True and False Positive Fraction values (SE and 1-SP) corresponding to each of the ODs highlighted in this figure are shown in Table 3.

4. Discussion and conclusions

The goal of this analysis was to evaluate the performance of the G-IFN test (SE, SP, and LR measures of association) under field conditions and the effect of different cut-off values on test accuracy. Several factors added value and originality to this study: (1) the data used in this analysis reflect how the test performs in the field where limitations in collection, shipping, and handling of samples exist, in contrast to how the test performs under controlled experimental conditions, (2) the data used to assess test SE reflect the performance of the test applied to animals with natural bTB infection in the low prevalence conditions characteristic of the U.S. cattle population, (3) the data used to calculate test SP reflect the performance of the test in cattle distributed throughout the United States having minimal or no risk of bTB infection, and lastly (4) this evaluation provides feedback on test performance to field veterinarians, animal health technicians, and DTEs who perform the test and use results to manage individual animals or herds.

The analysis of test SE was based on a limited number of confirmed bTB infected animals ($n = 87$) from database #1. Our results indicated that the SE of the G-IFN for a cut-off value of ≥ 0.1 lies between 76.1% and 91.6% with a most likely value of 83.9%. Similar SE values, between 85% and 90%, were reported in a study conducted by Ryan et al. (2000) in which the G-IFN was used under comparable conditions as in the United States (8 to 28 days after the injection of tuberculin). Other published studies where the G-IFN test was used as a primary test (i.e., not immediately after an intradermal injection of bovine PPD) also reported SE values ranging from 80% to 85% for the G-IFN test (Wood et al., 1991, 1992; Neill et al., 1994; Whipple et al., 1995). Test SE is expected to decrease with a cut-off value higher than 0.1. However, the small number of infected animals in our data limited the evaluation of SE for cut-off values other than ≥0.1 (value recommended by the manufacturer).

The literature reports a median SP value of 96.6% with a range between 85% and 99.6% (de la Rua-Domenech et al., 2006). The variability is due to differences in cattle populations, cut off values adopted to classify test results, test kit lots and reagents, and the gold standard used for classification of bTB infection status. In our analysis, the SP also depended on the cut-off value chosen to classify test results. A cut-off value ≥0.1, recommended by the manufacturer, and generally used when the G-IFN is applied in parallel with the CFT, resulted in a SP of approximately 90%. This SP value is within the range of previously published SP estimates (de la Rua-Domenech et al., 2006). On the other hand, cut-off values ≥0.3 and ≥0.5, often used in series testing with the CFT, yielded SP estimates of 97% and 98.6%, respectively. The estimation of test SP was based on testing data from cattle assumed to be bTB-free. This assumption based on the criteria used in the selection of the data, which excluded animals and premises associated with bTB investigations, outbreaks, and traces.

Test accuracy is dependent on the gold standard used to classify animal disease status. The criteria used in this study were largely based on positive results to culture or histopathology combined with PCR from tissue samples collected during postmortem examination of the animals. This gold standard, however, is not perfect. It is possible that animals in early stages if infection were misclassified as not infected when, in fact, they were infected. Consequently, the SE of the G-IFN in a herd where most cattle is in incipient stages of infection may be lower than the SE reported in this study. This bias is inevitable because it is rooted in the imperfection of the current gold standard and likely to be present in most studies addressing the accuracy of tests for bTB.

The specificity and sensitivity estimates obtained with the G-IFN test in our study are also comparable to those obtained with the comparative cervical skin test when used as a supplementary test for the CFT. The literature reports a SE for CCT lying between 52% and 100% with median values of approximately 80% when test interpretation excludes suspect animals from the "reactor"[2] category, and 93.5% when suspects are included in the "reactor" group (de la Rua-Domenech et al., 2006). The advantage of using the G-IFN over the CCT is that infected animals can be removed from herds rapidly, thereby reducing the risk of spread of infection. Herd owners and managers also appreciate the faster resolution of suspect cases and the decreased need for animal handling that characterizes the G-IFN versus the intradermal skin test.

The SE and SP evaluated in this analysis referred exclusively to the SE and SP of the G-IFN test applied between 3 and 30 days after a CFT (as indicated in the 2005 UM&R) under field conditions, and not to the SE and SP of the overall testing protocol. However, the SE and SP estimates reported in this document could be valuable inputs for modeling the overall SE and SP of parallel or series testing schemas (for CFT and G-IFN) in test and slaughter protocols.

The distribution of OD value results among infected and non-infected animals was mostly clustered around central values, but there were a few dispersed OD value results in both infected and non-infected cattle groups. Despite some misclassification of test results, both the likelihood ratios and the AUC indicated that the accuracy of the G-IFN test is moderate to high.

The LRs also provide a way to update the veterinarian's pre-test probability of bTB into a post-test probability given a test result. The pre-test probability of bTB infection is generally based on the veterinarian's knowledge about the clinical history of the animals, the risk of disease, the disease prevalence in the herd of origin, and previous individual and herd test results. According to guidelines cited by Gardner and Greiner (2006), the values of the LRs provided in this analysis suggest that a G-IFN test result ≥0.1 adds substantive evidence that an animal is infected (LRP = 9.0), while test results <0.1 predict a moderate decrease in post-test probability from the pre-test probability of bTB (LRN = 0.18). The LR, SE, and SP estimates from this study could be used to estimate case-specific post-test probabilities of bTB for different hypothetical prior probability scenarios using the odds form of Bayes' theorem (Gardner and Greiner, 2006).

In the medical field, ROC analysis has offered a visual approach to the analysis and selection of diagnostic systems as classifiers based on performance (Swets, 1998). In this paper, ROC analysis was used only to show the tradeoff between SE and SP at different cut-off values, not to report the specific values of SE at different cut-offs, since those specific values will be strongly influenced by the number of records and the actual OD values in the infected cattle group. Note that the SE estimate at a cut-off value of ≥0.1 from the ROC curve differed from the SE calculated by the traditional method (number of test positive/number of truly infected). This difference is based on the sample size used for calculation of the estimate (72 animals in ROC curve analysis versus 87 animals in the traditional calculation method). The SE reported in this study is based on the traditional method because the additional 15 animals add more power to the calculation of the SE estimate.

The selection of an optimal cut-off value depends on several factors, such as the prevalence of disease and costs incurred with false positive and false negative test results. The costs incurred, however, will depend on the rate of

[2] In this study "reactors" represented all those animals that responded to the CCT.

false positives and negatives, which depend in turn on the testing schema that is applied to the herd; that is, whether the G-IFN test is used with CFT to increase the SE or in lieu of CCT to increase the SP of the testing protocol. Precision of the SE and SP estimated by the ROC analysis at each cut-off value depends on the number of bTB infected and non-infected animals included in the data set. This analysis, however, confirms that lower cut-off values, such as 0.1, are most suitable to parallel testing where maximizing SE is the goal, while cut-off values ranging between 0.4 and 0.6 provide the high SP desired for series testing (or other testing protocols) maximizing SP.

Limitations of the current analysis included the lack of standardization and accessibility of data; data were located in multiple formats and in multiple databases. The selection criteria requiring test results with complete diagnostic follow-up and quantitative recording of G-IFN OD values resulted in a sample size too small to reliably evaluate the impact of factors such as age, geographic location, seasonality, the presence of coexisting infections with closely related mycobacteria (i.e., paratuberculosis or other mycobacteria spp.), or cattle type on the performance of the G-IFN test under field conditions.

This study evaluated how the G-IFN test performed under field conditions in the United States between 2005 and 2009. The study concludes that the G-IFN performed with high accuracy in the field, yielding SE and SP estimates comparable to those reported in previous evaluations (Ryan et al., 2000; Ameni et al., 2000; de la Rua-Domenech et al., 2006; Gormley et al., 2006). In addition, this study showed that by increasing the cut off value, the specificity of the G-IFN test gets closer to that of the CCT, which is also used as a supplementary test to the CFT.

References

Ameni, G., Miorner, H., Roger, F., Tibbo, M., 2000. Comparison between comparative tuberculin and gamma interferon tests for the diagnosis of bovine tuberculosis in Ethiopia Trop. Anim. Health Prod. 32 (5), 267–276.

Cagiola, M., Feliziani, F., Severi, G., Pasquali, P., Rutili, D., 2004. Analysis of possible factors affecting the specificity of the gamma interferon test in tuberculosis-free cattle herds. Clin. Diagn. Lab. Immunol. 11, 952–956.

de la Rua-Domenech, R., Goodchild, A.T., Vordermeier, H.M., Hewinson, R.G., Christiansen, K.H., Clifton-Hadley, R.S., 2006. Ante mortem diagnosis of tuberculosis in cattle: a review of the tuberculin test. G-IFN assay and other ancillary diagnostic techniques. Res. Vet. Sci. 81, 190–210.

Gardner, I.A., Greiner, M., 2006. Receiver-operating characteristic curves and likelihood ratios: improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests. Vet. Clin. Pathol., 358–417.

Gormley, E., Doyle, M.B., Fitzsimons, T., McGill, K., Collins, J.D., 2006. Diagnosis of Mycobacterium bovis infection in cattle by use of the gamma-interferon (Bovigam®) assay. Vet. Microbiol. 112 (2–4 SPEC. ISS), 171–179.

Neill, S.D., Cassidy, J., Hanna, J., Mackie, D.P., Pollock, M., Clements, A., Walton, E., Bryson, D.G., 1994. Detection of *Mycobacterium bovis* infection in skin test-negative cattle with an assay for bovine interferon-gamma. Vet. Rec. 135 (a), 134–135.

Palmer, M.V., Waters, W.R., Thacker, T.C., Greenwald, R., Esfandiari, J., Lyaschenko, K.P., 2006. Effects of different tuberculin skin-testing regimens on Gamma Interferon and antibody responses in cattle experimentally infected with Mycobacterium bovis. Clin. Vaccine Immunol. 13, 387–394.

Rothel, J.S., Jones, S.L., Corner, L.A., Cox, J.C., Wood, P.R., 1990. Aandwich enzyme immunoassay for bovine interferon and its use for the detection of tuberculosis in cattle. Aust. Vet. J. 67, 134–137.

Ryan, T.J., Buddle, B.M., DeLisle, G.W., 2000. An evaluation of the gamma interferon test for detecting bovine tuberculosis in cattle 8 to 28 days after tuberculin skin testing. Res. Vet. Sci. 69, 57–61.

Swets, J.A., 1998. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

United States Animal Health Association (USAHA), 2002. Report of the Committee on Tuberculosis. In: Proceedings of the USAHA Annual Meeting , St. Louis, MO, October, pp. 590–610.

USDA-APHIS-VS-Centers for Epidemiology and Animal Health (CEAH), 1995. Benefit–Cost and Performance Analysis of the Gamma Interferon test for Bovine Tuberculosis. USDA-APHIS-VS-Centers for Epidemiology and Animal Health (CEAH), Fort Collins, CO.

USDA-APHIS, 2005. Bovine Tuberculosis Eradication, Uniform Methods and Rules Effective. APHIS, 91-45-011.

Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Nabors, R.S., Payeur, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995. Comparison of the sensitivity of the caudal fold test and a commercial γ-interferon assay for diagnosis of bovine tuberculosis. Am. J. Vet. Res. 56, 415–419.

Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., McCormick, B.S., Francis, B.R., Creeper, J., Tweedle, N.E., 1991. Field comparison of the interferon gamma assay and the intradermal tuberculin test for the diagnosis of bovine tuberculosis. Aust. Vet. J. 68, 286–290.

Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., McCormick, B.S., Francis, B., Melville, L., Small, K., De Witte, K., Tolson, J., Ryan, T.J., deLisle, G.W., Cox, J.C., Jones, S.L., 1992. A field evaluation of serological and cellular diagnostic tests for bovine tuberculosis. Vet. Microbiol. 31, 71–79.